

Example 1: From Between-Person to Within-Person Models for Longitudinal Data (complete data, syntax, and output available for SAS, SPSS, and STATA electronically)

The models for this example come from Hoffman (2014) chapter 3. We will be examining the extent to which a learning achievement outcome can be predicted from group (control as the reference vs. treatment) and time (pre-test as the reference vs. post-test) in a sample of 50 children. We predict an interaction, such that the learning outcome should be higher at post-test than at pre-test, with a greater difference in the treatment group.

SAS Syntax for Data Manipulation:

```
* Location for files to be saved - CHANGE THIS TO YOUR DIRECTORY;
%LET example= C:\Dropbox\14_ICPSR\Download\SAS;
LIBNAME example "&example.";

* Open SAS stacked version of ANOVA data into work (temporary) library;
* Centering predictors for analysis;
DATA Example1; SET example.Example1;
    time1 = time - 1; * Time was coded 1,2;
    treat = group - 1; * Group was coded 1,2;
    LABEL time1 = "Time (0=pre-test, 1=post-test)"
           treat = "Treatment Group (0=control, 1=treatment)";
RUN;

TITLE1 "Means by group and time for learning outcome";
PROC MEANS MEAN STDERR MIN MAX DATA=Example1;
    CLASS group time; * Get means per group and time;
    WAYS 0 1 2; * Overall, marginal, cell means;
    VAR outcome; * List variables to describe;
RUN; TITLE1;
```

SPSS Syntax for Data Manipulation:

```
* Define locations of files used in examples -- CHANGE THIS.
FILE HANDLE example /NAME = "C:\Dropbox\14_ICPSR\Download\SPSS".

* SPSS code to center predictors for analysis.
GET FILE = "example/Example1.sav".
DATASET NAME example1 WINDOW=FRONT.
COMPUTE time1 = time - 1.
COMPUTE treat = group - 1.
VARIABLE LABELS time1 "Time (0=pre-test 1=post-test)"
                treat "Treatment Group (0=control, 1=treatment)".
EXECUTE.

TITLE "Means by group and time for learning outcome".
SUMMARIZE
    /TABLES = outcome BY group BY time
    /FORMAT = NOLIST TOTAL
    /CELLS = COUNT MEAN SEMEAN MIN MAX.
```

STATA Syntax for Data Manipulation:

```
* Open file used in example -- CHANGE THIS
use "C:\Dropbox\14_ICPSR\Download\STATA\Example1.dta", clear

* STATA code to center predictors for analysis
gen time1 = time - 1
gen treat = group - 1
label variable time1 "Time (0=pre-test 1=post-test)"
label variable treat "Treatment Group (0=control, 1=treatment)"

* Means by group and time for learning outcome
tabulate group time, summarize(outcome)
```

SAS Output:**Grand mean for learning outcome**

Mean	Std Error	Minimum	Maximum
53.3395829	0.6351006	37.5335041	68.6163028

Marginal means by group for learning outcome

Treatment Group (1=control, 2=treatment)		N Obs	Mean	Std Error	Minimum	Maximum
1	50	51.9879804	0.8943692	37.5335041	67.1060321	
2	50	54.6911855	0.8691455	40.5321932	68.6163028	

Marginal means by time for learning outcome

Time (1=pre-test 2=post-test)	N Obs	Mean	Std Error	Minimum	Maximum
1	50	49.9177687	0.7297690	37.5335041	62.1309134
2	50	56.7613972	0.7870204	44.5615778	68.6163028

Cell means by group and time for learning outcome

Treatment Group (1=control, 2=treatment)	Time (1=pre-test 2=post-test)	N Obs	Mean	Std Error	Minimum	Maximum
1	1	25	49.0767977	1.1370576	37.5335041	59.5504810
	2	25	54.8991630	1.1256529	44.5615778	67.1060321
2	1	25	50.7587396	0.9070808	40.5321932	62.1309134
	2	25	58.6236314	0.9864754	47.4303443	68.6163028

3.1: Between-Person Empty Means Model $y_{ti} = \beta_0 + e_{ti}$

```
TITLE1 "SAS Between-Person Empty Means Model via MIXED";
PROC MIXED DATA=Example1 NOITPRINT NOCLPRINT COVTEST IC METHOD=REML;
  CLASS PersonID time;
  MODEL outcome = / SOLUTION DDFM=BW;
  REPEATED time / R RCORR TYPE=VC SUBJECT=PersonID;
RUN; TITLE1;
```

METHOD = ML or REML (default)
 CLASS = categorical predictors, nesting
 MODEL dv = fixed effects / print solution
 REPEATED = residuals in **R** matrix

```
TITLE "SPSS Between-Person Empty Means Model via MIXED".
MIXED outcome BY PersonID time
/METHOD = REML
/PRINT = SOLUTION TESTCOV R
/FIXED =
/REPEATED = time | SUBJECT(PersonID) COVTYPE(ID).
```

MIXED dv BY categorical predictors
 WITH continuous predictors
 /METHOD = REML or ML
 /PRINT = regression solution
 /FIXED = predictors for means model
 /REPEATED = residuals in **R** matrix

```
* STATA Between-Person Empty Means Model via mixed
mixed outcome , || PersonID: , noconstant ///
  variance reml residuals(independent,t(time)),
  estat wcorrelation, covariance, // R matrix
  estat wcorrelation, // RCORR matrix
```

DV = outcome, random part after ||
 Level 2 ID is PersonID, random intercept by
 default, so noconstant removes it
 Print variances instead of SD, use REML
 residuals(independent) → type of R matrix by time

SAS Output:

Dimensions

Covariance Parameters	1
Columns in X	1
Columns in Z	0
Subjects	50
Max Obs Per Subject	2

This table tells you how many parameters are in your model for the means (“columns in x”, the fixed effects, or 1 fixed intercept here) and in your model for the variances (“covariance parameters”, or 1 residual variance here). It also tells you how many observations were read per subject, as defined by SUBJECT= on the REPEATED line.

Number of Observations

Number of Observations Read	100
Number of Observations Used	100
Number of Observations Not Used	0

Estimated R Matrix

for PersonID 1

Row	Col1	Col2
1	40.3353	
2		40.3353

$$\begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}$$

This **R** matrix says that the learning outcome has equal residual variance at pre-test (row 1) as at post-test (row 2), with no covariance between occasions.

Estimated R Correlation

Matrix for PersonID 1

Row	Col1	Col2
1	1.0000	
2		1.0000

Covariance Parameter Estimates

Cov	Subject	Estimate	Standard Error	Z	Pr > Z
time	PersonID	40.3353	5.7330	7.04	<.0001

This is the estimate of the residual variance σ_e^2 . It is labeled “time” because that is how the **R** matrix is structured via the REPEATED line.

Fit Statistics

-2 Res Log Likelihood	651.6
AIC (smaller is better)	653.6
AICC (smaller is better)	653.6
BIC (smaller is better)	655.5

The -2LL is the deviance index of model misfit. The other indices are “information criteria” that penalize misfit for parsimony (as we’ll see later).

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
0	0.00	1.0000

This “null model” LRT examines the need for any random effects variances and covariances. Because we don’t have any (yet), df = 0.

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	53.3396	0.6351	49	83.99	<.0001

This is the estimate of the fixed intercept β_0 .

A note about differences between program output:

SPSS output is very similar to SAS output, but RCORR (and later, GCORR and VCORR) is not available.

STATA output is also similar, with one big difference: Deviance (-2LL) is not provided directly. Instead, the model log-likelihood (LL) is provided instead (and information criteria are not given by default). In addition, the fixed effects are presented first, with the fixed intercept labeled “constant” instead.

3.2: Within-Person Empty Means Model $y_{ti} = \beta_0 + U_{0i} + e_{ti}$

```

TITLE1 "SAS Within-Person Empty Means Model via MIXED";
PROC MIXED DATA=Example1 NOITPRINT NOCLPRINT COVTEST IC METHOD=REML;
  CLASS PersonID time;
  MODEL outcome = / SOLUTION DDFM=BW;
  REPEATED time / R RCORR TYPE=CS SUBJECT=PersonID;
RUN; TITLE1;

TITLE "SPSS Within-Person Empty Means Model via MIXED".
MIXED outcome BY PersonID time
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV R
  /FIXED =
  /REPEATED = time | SUBJECT(PersonID) COVTYPE(CS).

* STATA Within-Person Empty Means Model via mixed
mixed outcome , || PersonID: , noconstant ///
  variance reml residuals(exchangeable,t(time)),
  estat wcorrelation, covariance, // R matrix
  estat wcorrelation // RCORR matrix

```

The model adds a distinction between BP and WP variability (and a covariance between occasions) but using an R matrix with a compound symmetry structure ($U_{0i} + e_{ti}$).

This is accomplished via TYPE=CS in SAS, COVTYPE(CS) in SPSS, and residuals(exchangeable) in STATA.

SAS Output:

	Dimensions
Covariance Parameters	2
Columns in X	1
Columns in Z	0
Subjects	50
Max Obs Per Subject	2

We still have 1 fixed effect, the fixed intercept, but now the model for the variance includes random intercept variance and residual variance.

Estimated R Matrix
for PersonID 1

Row	Col1	Col2
1	40.4590	12.2526
2	12.2526	40.4590

$$\begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

Estimated R Correlation
Matrix for PersonID 1

Row	Col1	Col2
1	1.0000	0.3028
2	0.3028	1.0000

$$\begin{bmatrix} 1 & \text{ICC} \\ \text{ICC} & 1 \end{bmatrix}$$

This **R** matrix now says that the learning outcome has equal residual variance at pre-test (row 1) as at post-test (row 2), with a covariance between occasions due only to the random intercept variance.

The intraclass correlation can then be computed as:

$$\text{ICC} = \frac{12.2526}{12.2526 + 28.2064} = .30$$

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
CS	PersonID	12.2526	6.0256	2.03	0.0420
Residual		28.2064	5.6413	5.00	<.0001

CS = Random Intercept Variance $\tau_{u_0}^2$
Residual = Residual Variance σ_e^2

Fit Statistics

-2 Res Log Likelihood	646.8
AIC (smaller is better)	650.8
AICC (smaller is better)	650.9
BIC (smaller is better)	654.6

Now we have a random intercept variance, so $df=1$. This is the model comparison of BP vs. WP empty means models via a likelihood ratio test (LRT, or deviance difference test): $651.6 - 646.8 = 4.77$. So this means the WP version fits better, and that the ICC = .30 is significantly > 0.

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	4.77	0.0289

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	53.3396	0.7260	49	73.47	<.0001

This is still the estimate of the fixed intercept β_0 , but note the SE differs.

3.7 (top): Between-Person Conditional Model

Model for outcome: $y_{ti} = \beta_0 + \beta_1(\text{Time}_{ti}) + \beta_2(\text{Group}_i) + \beta_3(\text{Time}_{ti})(\text{Group}_i) + e_{ti}$

Simple slope for time: $[\beta_1 + \beta_3(\text{Group}_i)](\text{Time}_{ti})$

Simple slope for group: $[\beta_2 + \beta_3(\text{Time}_{ti})](\text{Group}_i)$

```
TITLE1 "SAS Between-Person Conditional (Predictor) Model via MIXED";
TITLE2 "Manually dummy coding group and time";
PROC MIXED DATA=Example1 NOITPRINT NOCLPRINT COVTEST IC METHOD=REML;
  CLASS PersonID time;
  MODEL outcome =time1 treat time1*treat / SOLUTION DDFM=BW;
  REPEATED time / R RCORR TYPE=VC SUBJECT=PersonID;
* Request cell means;
ESTIMATE "Mean: Control Group at Pre-Test"      intercept 1 time1 0 treat 0 time1*treat 0;
ESTIMATE "Mean: Control Group at Post-Test"      intercept 1 time1 1 treat 0 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Pre-Test"     intercept 1 time1 0 treat 1 time1*treat 0;
ESTIMATE "Mean: Treatment Group at Post-Test"    intercept 1 time1 1 treat 1 time1*treat 1;
* Request simple slopes;
ESTIMATE "Time Effect for Control Group"         time1 1 time1*treat 0;
ESTIMATE "Time Effect for Treatment Group"       time1 1 time1*treat 1;
ESTIMATE "Group Effect at Pre-Test"              treat 1 time1*treat 0;
ESTIMATE "Group Effect at Post-Test"             treat 1 time1*treat 1;
RUN; TITLE1; TITLE2;
```

```
TITLE "SPSS Between-Person Conditional Model via MIXED".
MIXED outcome BY PersonID time WITH time1 treat
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV R
  /FIXED = time1 treat time1*treat
  /REPEATED = time | SUBJECT(PersonID) COVTYPE(ID)
  /TEST "Mean: Control Group at Pre-Test"      intercept 1 time1 0 treat 0 time1*treat 0
  /TEST "Mean: Control Group at Post-Test"     intercept 1 time1 1 treat 0 time1*treat 0
  /TEST "Mean: Treatment Group at Pre-Test"    intercept 1 time1 0 treat 1 time1*treat 0
  /TEST "Mean: Treatment Group at Post-Test"   intercept 1 time1 1 treat 1 time1*treat 1
  /TEST "Time Effect for Control Group"        time1 1 time1*treat 0
  /TEST "Time Effect for Treatment Group"      time1 1 time1*treat 1
  /TEST "Group Effect at Pre-Test"             treat 1 time1*treat 0
  /TEST "Group Effect at Post-Test"            treat 1 time1*treat 1.
```

```
* STATA Between-Person Conditional Model via mixed
mixed outcome c.time1 c.treat c.time1#c.treat, ///
  || PersonID: , noconstant ///
  variance reml residuals(independent,t(time)),
  estat wcorrelation, covariance, // R matrix
  estat wcorrelation, // RCORR matrix
* Request cell means
margins, at(c.time1=(0 1) c.treat=(0 1)) vsquish
* Request simple slopes
lincom 1*c.time + 0*c.time#c.treat // time effect for control
lincom 1*c.time + 1*c.time#c.treat // time effect for treat
lincom 1*c.treat + 0*c.time#c.treat // group effect for pre-test
lincom 1*c.treat + 1*c.time#c.treat // group effect for post-test
```

SAS Output:

	Dimensions
Covariance Parameters	1
Columns in X	4
Columns in Z	0
Subjects	50
Max Obs Per Subject	2

Now we have 4 parameters in the model for the means and 1 parameter in the model for the variance (just σ_e^2).

Estimated R Matrix
for PersonID 1

Row	Col1	Col2
1	27.2245	
2		27.2245

$$\begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}$$

This **R** matrix says that, after allowing differences due to time*group, that the learning outcome has equal residual variance at pre-test (row 1) as at post-test (row 2), with no covariance between occasions.

Estimated R Correlation
Matrix for PersonID 1

Row	Col1	Col2
1	1.0000	
2		1.0000

Covariance Parameter Estimates

Cov	Subject	Estimate	Standard Error	Z	Pr > Z
time	PersonID	27.2245	3.9295	6.93	<.0001

This is the estimate of the residual variance σ_e^2 . It is labeled "time" because that is how the R matrix is structured via the REPEATED line.

Fit Statistics

-2 Res Log Likelihood	602.5
AIC (smaller is better)	604.5
AICC (smaller is better)	604.5
BIC (smaller is better)	606.4

BP Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t	
Intercept	49.0768	1.0435	48	47.03	<.0001	beta0
time1	5.8224	1.4758	48	3.95	0.0003	beta1
treat	1.6819	1.4758	48	1.14	0.2601	beta2
time1*treat	2.0425	2.0871	48	0.98	0.3327	beta3

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	
Mean: Control Group at Pre-Test	49.0768	1.0435	48	47.03	<.0001	
Mean: Control Group at Post-Test	54.8992	1.0435	48	52.61	<.0001	
Mean: Treatment Group at Pre-Test	50.7587	1.0435	48	48.64	<.0001	
Mean: Treatment Group at Post-Test	58.6236	1.0435	48	56.18	<.0001	
Time Effect for Control Group	5.8224	1.4758	48	3.95	0.0003	beta1
Time Effect for Treatment Group	7.8649	1.4758	48	5.33	<.0001	beta1+beta3
Group Effect at Pre-Test	1.6819	1.4758	48	1.14	0.2601	beta2
Group Effect at Post-Test	3.7245	1.4758	48	2.52	0.0150	beta2+beta3

Means (SE)	Pre-Test	Post-Test	Marginal
Control	49.08 (1.14)	54.90 (1.13)	51.99 (0.89)
Treatment	50.76 (0.91)	58.62 (0.99)	54.70 (0.87)
Marginal	49.92 (0.73)	56.76 (0.79)	53.34 (0.64)

These results assume independent observations... what happens if that's not the case?

3.7 (bottom): Within-Person Conditional Model

Model for outcome: $y_{ti} = \beta_0 + \beta_1(\text{Time}_{ti}) + \beta_2(\text{Group}_i) + \beta_3(\text{Time}_{ti})(\text{Group}_i) + U_{0i} + e_{ti}$

Simple slope for time: $[\beta_1 + \beta_3(\text{Group}_i)](\text{Time}_{ti})$

Simple slope for group: $[\beta_2 + \beta_3(\text{Time}_{ti})](\text{Group}_i)$

The code is exactly the same, except for a compound symmetry R matrix instead of variance components:

SAS: `REPEATED time / R RCORR TYPE=CS SUBJECT=PersonID;`

SPSS: `/REPEATED = time | SUBJECT(PersonID) COVTYPE(CS)`

STATA: `residuals(exchangeable,t(time))`

SAS Output:

Dimensions
Covariance Parameters 2
Columns in X 4
Columns in Z 0
Subjects 50
Max Obs Per Subject 2

We still have 4 parameters in the model for the means, but now we have 2 parameters in the model for the variance ($\tau_{U_0}^2$ and σ_e^2).

Estimated R Matrix
for PersonID 1

Row	Col1	Col2
1	27.2245	22.7794
2	22.7794	27.2245

$$\begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

This **R** matrix now says that, after allowing differences due to time*group, the learning outcome has equal residual variance at pre-test (row 1) as at post-test (row 2), with a covariance between occasions due only to the random intercept variance.

Estimated R Correlation
Matrix for PersonID 1

Row	Col1	Col2
1	1.0000	0.8367
2	0.8367	1.0000

$$\begin{bmatrix} 1 & \text{ICC} \\ \text{ICC} & 1 \end{bmatrix}$$

The conditional intraclass correlation can be computed as:

$$\text{ICC} = \frac{22.7794}{22.7794 + 4.4451} = .84$$

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z	Pr > Z
CS	PersonID	22.7794	5.1236	4.45	<.0001
Residual		4.4451	0.9073	4.90	<.0001

CS = Random Intercept Variance $\tau_{U_0}^2$
Residual = Residual Variance σ_e^2

Fit Statistics

-2 Res Log Likelihood	544.7
AIC (smaller is better)	548.7
AICC (smaller is better)	548.8
BIC (smaller is better)	552.5

Now we have a random intercept variance, so $df=1$. This is the model comparison of BP vs. WP conditional models via a likelihood ratio test (LRT, or deviance difference test): $602.5 - 544.7 = 57.81$. So this means the WP version fits better, and that the $\text{ICC} = .84$ is significantly > 0 .

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	57.81	<.0001

This also means the results from the WP model with respect to the fixed effects should be more accurate than those of the BP model... let's see how they differ.

WP Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	49.0768	1.0435	48	47.03	<.0001 beta0
time1	5.8224	0.5963	48	9.76	<.0001 beta1
treat	1.6819	1.4758	48	1.14	0.2601 beta2
time1*treat	2.0425	0.8433	48	2.42	0.0193 beta3

WP Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
Mean: Control Group at Pre-Test	49.0768	1.0435	48	47.03	<.0001
Mean: Control Group at Post-Test	54.8992	1.0435	48	52.61	<.0001
Mean: Treatment Group at Pre-Test	50.7587	1.0435	48	48.64	<.0001
Mean: Treatment Group at Post-Test	58.6236	1.0435	48	56.18	<.0001
Time Effect for Control Group	5.8224	0.5963	48	9.76	<.0001 beta1
Time Effect for Treatment Group	7.8649	0.5963	48	13.19	<.0001 beta1+beta3
Group Effect at Pre-Test	1.6819	1.4758	48	1.14	0.2601 beta2
Group Effect at Post-Test	3.7245	1.4758	48	2.52	0.0150 beta2+beta3

Which results differ from those of the BP version of the model (repeated below), and why?

BP Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	49.0768	1.0435	48	47.03	<.0001 beta0
time1	5.8224	1.4758	48	3.95	0.0003 beta1
treat	1.6819	1.4758	48	1.14	0.2601 beta2
time1*treat	2.0425	2.0871	48	0.98	0.3327 beta3

BP Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
Mean: Control Group at Pre-Test	49.0768	1.0435	48	47.03	<.0001
Mean: Control Group at Post-Test	54.8992	1.0435	48	52.61	<.0001
Mean: Treatment Group at Pre-Test	50.7587	1.0435	48	48.64	<.0001
Mean: Treatment Group at Post-Test	58.6236	1.0435	48	56.18	<.0001
Time Effect for Control Group	5.8224	1.4758	48	3.95	0.0003 beta1
Time Effect for Treatment Group	7.8649	1.4758	48	5.33	<.0001 beta1+beta3
Group Effect at Pre-Test	1.6819	1.4758	48	1.14	0.2601 beta2
Group Effect at Post-Test	3.7245	1.4758	48	2.52	0.0150 beta2+beta3

If you think about these data as derived from a three-way crossed design of time by group by person... What other terms that could possibly be included are missing? Are they *really* missing?

Model for outcome: $y_{ti} = \beta_0 + \beta_1(\text{Time}_{ti}) + \beta_2(\text{Group}_i) + \beta_3(\text{Time}_{ti})(\text{Group}_i) + U_{0i} + e_{ti}$